Simple GMM Estimation of the Semi-Strong GARCH(1,1) Model

Prono Todd

Commodity Futures Trading Commission

January 2010

Simple GMM Estimation of the Semi-Strong GARCH(1,1) Model

Todd Prono[1]
Commodity Futures Trading Commission

First Version: November 2009
This Version: January 2010

Abstract

Efficient GMM estimation of the semi-strong GARCH(1,1) model requires simultaneous estimation of the conditional third and fourth moments. This paper proposes a simple alternative to efficient GMM based upon the unconditional skewness of residuals and the autocovariances of squared residuals. An advantage of this simple alternative is that neither the third nor the fourth conditional moment needs to be estimated. A second advantage is that linear estimators apply to all of the parameters in the model, making estimation straightforward in practice. The proposed estimators are IV-like with potentially many instruments. Sequential estimation involves TSLS in a first step followed by linear GMM. Simultaneous estimation involves either two-step GMM or CUE. A Monte Carlo study of the proposed estimators is included.

Keywords: GARCH, Time Series Heteroskedasticity, GMM, CUE, Many Moments, Conditional Moment Restrictions, Consistency, Robust Statistics. JEL codes: C22, C53, G12.

[1] Corresponding Author: Todd Prono, Commodity Futures Trading Commission, Office of the Chief Economist, 1155 21st, N.W., Washington, DC 20581. (202) 418-5460, tprono@cftc.gov.

1. Introduction

Despite a plethora of alternative volatility models intended to capture certain "stylized facts" of financial time series, the standard GARCH(1,1) model of Bollerslev (1986) remains the workhorse of conditional heteroskedasticity (CH) modeling in financial economics. By far, the most common estimator for this model is QML. While in an IV context, efficient GMM estimation is also possible, the instruments required are nonlinear functions of the third and fourth conditional moments as well as derivatives of the conditional variance function. This paper develops simple GMM estimators for the GARCH(1,1) model also with an IV interpretation, but where the instruments are only a small (relative to the sample size) collection of past residuals and squared residuals. The advantage of these simple estimators over efficient GMM is that the conditional third and fourth moments do not need to be estimated. The advantage over QML is that estimation of the ARCH and GARCH parameters can be conducted with linear estimators.

Weiss (1986) first demonstrates the CAN properties of the QMLE for ARCH models. Lumsdaine (1996) relaxes some of the conditions from Weiss in her study of the GARCH(1,1) model, but continues to assume that the model's standardized residuals are iid. It is well known that financial return data often exhibit non-zero skewness and excess kurtosis. Works by such authors as Hansen (1994) and Harvey and Siddique (1999, 2000), find this skewness and kurtosis to be time-varying. These findings do not square with the notion that conditional dependence be relegated to the first two moments. While Bollerslev and Wooldridge (1992), Lee and Hansen (1994), and Escanciano (2009) investigate the CAN properties of the QMLE minus the need for iid innovations (i.e., they study the asymptotic properties of the QML estimator for semi-strong GARCH processes; see Drost and Nijman 1993 for a definition), this estimator does not utilize any of the information contained in the higher moments.

As recognized by Bollerslev and Wooldridge (1992), the "results of Chamberlain (1982), Hansen (1982), White (1982), and Cragg (1983) can be extended to produce an instrumental variables estimator asymptotically more efficient than QMLE under nonnormality" (p. 5-6). Skoglund (2001) demonstrates this claim for the strong GARCH(1,1) model. The drawback of such an approach to semi-strong GARCH(1,1) estimation is the need to either parameterize or treat nonparametrically the conditional third and fourth moments. Weis (1986), Rich, Raymond and Butler (1991), and Guo and Phillips (2001) discuss GMM estimation of the ARCH(p) model given the existence of a finite fourth moment. Their results have the advantage of not requiring treatment of the third and fourth moment dynamics. However, their results do not extend to the GARCH(1,1) case because the autocovariances of squared residuals do not separately identify the ARCH and GARCH terms. This paper uses cross-moment covariances and squared residual autocovariances to identify the GARCH(1,1) model. The key to identification is nonzero skewness of the residuals. Consistency of the resulting estimator, therefore, only requires a finite third moment. Two-stage least squares can be used to estimate the ARCH parameter. Conditional on this estimate, the GARCH parameter can then be retrieved with linear GMM.

The remainder of this paper is organized as follows. Section 1.1 briefly discusses how the testing of a common model for pricing risky assets would benefit from the estimators proposed in this paper. Section 2 outlines the model's assumptions, states two lemmas that define a set of moment conditions and proposes a GMM estimator based upon these moment conditions. Section 3 establishes consistency of this estimator and a multi-step approach comprised entirely of linear estimators. A generalized IV-estimator for the ARCH(1) model is also proposed, and a method for calculating standard errors and conducting specification testing is discussed. Section 4 summarizes the results from Monte Carlo studies of the proposed estimators. Section 5 concludes.

1.1 A Conditional Asset Pricing Model

For the sequence $\left\{\left(r_{i,t},\ r_{m,t}\right), \quad i=1,\ldots,N;\ t=1,\ldots,T\right\}$, let $r_{i,t}$ and $r_{m,t}$ be the return on the ith risky asset and the return on the market for all risky assets, respectively, measured in excess of an observable risk free rate. Let J_{t-1} be the set of information observable to the econometrician at time $t-1$. Consider the following model for risky assets:

$$\begin{aligned} r_{i,t} &= \frac{E\left[u_{i,t}u_{m,t} \mid J_{t-1}\right]}{E\left[u_{m,t}^2 \mid J_{t-1}\right]} E\left[r_{m,t} \mid J_{t-1}\right] + u_{i,t} \\ r_{m,t} &= E\left[r_{m,t} \mid J_{t-1}\right] + u_{m,t} \end{aligned} \tag{1}$$

where $u_{i,t}$ and $u_{m,t}$ are both mean zero residuals conditional on J_{t-1}. Since $cov\left[r_{i,t},\ r_{m,t} \mid J_{t-1}\right] = E\left[u_{i,t}u_{m,t} \mid J_{t-1}\right]$, and $var\left[r_{m,t} \mid J_{t-1}\right] = E\left[u_{m,t}^2 \mid J_{t-1}\right]$, (1) is a statement of the conditional CAPM, where the conditional risk premium for the ith asset is a function of its conditional beta and the conditional risk premium for the market. A large literature centers around testing various specifications of (1).

Estimation of (1) requires specification of the conditional moments $E\left[r_{m,t} \mid J_{t-1}\right]$, $E\left[u_{i,t}u_{m,t} \mid J_{t-1}\right]$, and $E\left[u_{m,t}^2 \mid J_{t-1}\right]$. Usually, $E\left[r_{m,t} \mid J_{t-1}\right] = X_{t-1}'\delta$, where X_{t-1} is a vector of supposed forecasting instruments for risky assets. Mark (1988) and Bodurtha and Mark (1991) specify $E\left[u_{i,t}u_{m,t} \mid J_{t-1}\right]$ and $E\left[u_{m,t}^2 \mid J_{t-1}\right]$ as low order ARCH processes. As a result, the system in (1) can be estimated by GMM using X_{t-1} and a collection of lagged squared residuals and cross-products of residuals Z_{t-1} as instruments. Given the estimators developed in this paper, $E\left[u_{m,t}^2 \mid J_{t-1}\right]$ can be generalized to a GARCH(1,1) process and the system can be estimated in the same way by simply supplementing Z_{t-1} with lags of the residuals. Moreover, if $E\left[u_{i,t}^2 \mid J_{t-1}\right]$ is also considered to be GARCH(1,1), then so too can $E\left[u_{i,t}u_{m,t} \mid J_{t-1}\right]$ given the method for estimating restricted bivariate diagonal GARCH(1,1) processes discussed in Prono (2006). Such generalizations seem advantageous for characterizing the time variation in conditional betas, since the GARCH(1,1) specification tends to dominate its ARCH(1) counterpart in terms of in-sample fit and out-of-sample forecasting power (see, e.g., Hansen and Lunde 2005), and since the performance of (1) is often characterized in terms of a test of the overidentifying restrictions from the GMM objective function.

2. The Model, Assumptions, and Estimation

For the sequence $\{Y_t,\ t \in \mathbb{Z}\}$, define Ψ_{t-1} as the σ-field generated by $\left\{Y_{t-1},\ Y_{t-2}, \ldots\right\}$. Consider the model

$$E\left[Y_t \mid \Psi_{t-1}\right] = 0, \qquad E\left[Y_t^2 \mid \Psi_{t-1}\right] = h_t \tag{2}$$

where

$$h_t = \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 h_{t-1}. \tag{3}$$

In what follows, ω_0 denotes the true value, ω any one of a set of possible values, and $\widehat{\omega}$ an estimate. Parallel distinctions hold for all other parameter values. The model of (2) and (3) defines the semi-strong GARCH process of Drost and Nijman (1993). The usual assumptions regarding this model's standardized residuals (i.e., that they originate from some known parametric distribution and that they are iid) are not made.

Let $\sigma_0^2 = \frac{\omega_0}{1-(\alpha_0+\beta_0)}$, and define $\theta_0 = (\sigma_0^2,\ \alpha_0,\ \beta_0)'$. The usual parameter vector considered for the GARCH(1,1) model is $\vartheta_0 = (\omega_0,\ \alpha_0,\ \beta_0)'$. Consideration of θ_0, instead, has the advantage of guaranteeing that the unconditional variance implied by the model equals the sample variance. Such a feature is particularly attractive in the current context since moments-based estimators of (3) are being considered. The VTE method of Engle and Mezrich (1996), the asymptotic properties of which are developed by Francq, Horath, and Zakoian (2009), replies upon a similar reparameterization. Retrieval of $\widehat{\omega}$ is straightforward given $\widehat{\theta}$.

ASSUMPTION A1: *The true parameter vector $\theta_0 \in \Theta \subseteq \Re^3$ is in the interior of Θ, a compact parameter space. For any $\theta \in \Theta$, there exists a $\partial \in (0,\ \frac{1}{2})$ such that $\partial \leq \omega \leq W$, $\partial \leq \alpha \leq 1-\partial$, and $0 \leq \beta \leq 1-\partial$, where ∂ and W are given a priori.*

The restrictions on θ ensure that h_t is everywhere strictly positive and that $\alpha + \beta < 1$. As a consequence, $\{Y_t\}$ is covariance stationary following Theorem 1 of Bollerslev (1986), with $E\left[Y_t^2\right] = \sigma_0^2$. From Lumsdaine (1996), α is strictly positive because if $\alpha = 0$, then h_t is completely deterministic, in which case ω_0 and β_0 are not separately identified. Since $\beta \geq 0$, A1 nests the ARCH(1) model.

The mean-adjusted form of (3) is

$$\widetilde{h}_t = \alpha_0 \widetilde{Y}_{t-1}^2 + \beta_0 \widetilde{h}_{t-1}, \tag{4}$$

where $\widetilde{h}_t = h_t - \sigma_0^2$ and $\widetilde{Y}_t^2 = Y_t^2 - \sigma_0^2$. An implication of (4) is that

$$\widetilde{Y}_t^2 = \widetilde{h}_t + W_t, \tag{5}$$

where $E\left[W_t \mid \Psi_{t-1}\right] = 0$. Guo and Phillips (2001) consider an analogous specification to (5) in their development of an efficient IV estimator for the ARCH(p) model. Recursively substituting $\widetilde{h}_{t-\tau}$ into (4) for $\tau \geq 1$ produces

$$\widetilde{h}_t = \sum_{i=0}^{t-1} \alpha_0 \beta_0^i \widetilde{Y}_{t-1-i}^2 + \beta_0^t \widetilde{h}_0, \tag{6}$$

for some arbitrary constant $\widetilde{h}_0$. Using (6) to solve (5) forward from $t = 1$ setting $\widetilde{Y}_0^2 = 0$ produces

$$\widetilde{Y}_t^2 = W_t + \alpha_0 \sum_{i=1}^{t-1} \left(\alpha_0 + \beta_0\right)^{i-1} W_{t-i} + \beta_0 \left(\alpha_0 + \beta_0\right)^{t-1} \widetilde{h}_0, \tag{7}$$

which shows that the GARCH(1,1) model relates $\widetilde{Y}_t^2$ to weighted sum of current and past W_t. The instruments from Guo and Phillips (2001) are based on weighted sums of innovations similar to (7). Properties of $\{W_t\}$ are central in defining simple GMM estimators for (3) and are the subject of the following two assumptions.

ASSUMPTION A2: *(i)* $E\left[W_t Y_t\right] = \gamma_0 \neq 0 \ \forall \ t$. *(ii) The sequence* $\{W_t Y_t - \gamma_0\}$ *is an* L^1 *mixingale as defined in Andrews (1988) that is uniformly integrable. (iv) The sequences* $\left\{W_{t-l} Y_{t-k}\right\}$ *where* $k, l = 1, \ldots, K$ *and* $k \neq l$ *are uniformly integrable.*

Given (5) and an application of iterated expectations,

$$\begin{aligned} E\left[Y_t^3\right] &= E\left[\widetilde{Y}_t^2 Y_t\right] \\ &= E\left[\left(\widetilde{h}_t + W_t\right) Y_t\right] \\ &= E\left[W_t Y_t\right] \end{aligned} \tag{8}$$

Given A2(i), therefore, $\{Y_t\}$ is asymmetric with a stationary third moment. Seen through (8), A2(ii) imposes restrictions on the process governing $E\left[Y_t^3 \mid \Psi_{t-1}\right]$. L^1 mixingales exhibit weak temporal dependence that need not decay towards zero at any particular rate and that include certain infinite order moving average and autoregressive moving average processes. Given the functional form of (3), allowing the third moment to display similar dynamics seems natural. Moreover, Harvey and Siddique (1999) present empirical evidence from stock return data that the conditional third moment is autoregressive. Uniform integrability allows a weak LLN to apply to $\{W_t Y_t - \gamma_0\}$ and $\left\{W_{t-l} Y_{t-k}\right\}$ (See Lemma 3 in the Appendix). A sufficient condition for this result is that the given sequence be L^p bounded for some $p > 1$. According to Andrews (1988), however, "it is preferable to impose the uniform integrability assumption rather than an L^p bounded assumption because the former allows for more heterogeneity in the higher order moments of the rv's" (p. 3).

ASSUMPTION A3: *(i)* $E\left[W_t^2\right] = \lambda_0 \, \forall \, t$. *(ii) The sequences* $\left\{W_t W_{t-k}\right\}$ *are uniformly integrable. (iii) The sequence* $\{W_t^2 - \lambda_0\}$ *is an* L^1 *mixingale that is uniformly integrable.*

Suppose

$$Y_t = \sqrt{h_t}\epsilon_t, \tag{9}$$

where $\{\epsilon_t\}$ is iid with a mean of zero and a unit variance. Then A3(i) is equivalent to assuming that

$$(\kappa + 1)\,\alpha_0^2 + 2\alpha_0\beta_0 + \beta_0^2 < 1; \quad \kappa = E\left[\epsilon_t^4\right] - 1,$$

which is the necessary and sufficient condition for establishing existence of the fourth moment of $\{Y_t\}$ according to Theorem 1 of Zadrozny (2005).[2] A3(ii)-(iii) permit a weak LLN to apply to the sample autocovariances of $\{Y_t^2\}$. A3(iii) assumes that the same general type of process

[2] If $\{\epsilon_t\}$ is normally distributed, then this inequality follows from Theorem 2 of Bollerslev (1986).

that governs the third moment (see A2ii) also governs the fourth. This assumption is supported empirically by the results of Hansen (1994).

LEMMA 1. *Let Assumptions A1 and A2(i) hold for the model of (2) and (3). Then*

$$E\left[\widetilde{Y}_t^2 Y_{t-1}\right] = \alpha_0 E\left[W_t Y_t\right], \tag{10}$$

and

$$E\left[\widetilde{Y}_t^2 Y_{t-(k+1)}\right] = \left(\alpha_0 + \beta_0\right) E\left[\widetilde{Y}_t^2 Y_{t-k}\right] \tag{11}$$

for $k \geq 1$.

All proofs are given in the Appendix. Lemma 1 relates the covariance between Y_t^2 and Y_{t-k} to the third moment of Y_t. Lemma 1 of Guo and Phillips (2001) establishes an analogous result for the ARCH(p) model. From (10), α_0 is identified as

$$\alpha_0 = E\left[\widetilde{Y}_t^2 Y_{t-1}\right] / E\left[Y_t^3\right].$$

Let $\widetilde{Z}_{t-2} = \left[\widetilde{Y}_{t-2}^2 \cdots \widetilde{Y}_{t-K}^2\right]'$. From (11), β_0 is then identified as

$$\beta_0 = \left(E\left[\widetilde{Y}_t^2 \widetilde{Z}_{t-1}\right]' E\left[\widetilde{Y}_t^2 \widetilde{Z}_{t-1}\right]\right)^{-1} E\left[\widetilde{Y}_t^2 \widetilde{Z}_{t-1}\right]' E\left[\widetilde{Y}_t^2 \widetilde{Z}_{t-2}\right] - \alpha_0.$$

Lemma 1, therefore, provides a moments-based identification condition for the GARCH(1,1) model.

Newey and Steigerwald (1997) explore the effects of asymmetry on the identification of CH models using the QML estimator. This paper conducts a similar exploration for the GMM estimator. Newey and Steigerwald (1997) show that given asymmetry, there exist conditions under which the standard QML estimator for CH models is not identified. In contrast, this paper develops a simple GMM estimator that is not identified without such asymmetry.

LEMMA 2. *Given the model of (2) and (3),* $\{Y_t^2\}$ *is covariance stationary if and only if A1 and*

A3(i) hold. In this case,

$$E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-(k+1)}^2\right] = (\alpha_0+\beta_0)E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-(k)}^2\right] \tag{12}$$

for $k \geq 1$.

Mark (1988) as well as Rich, Raymond, and Butler (1991) estimate ARCH models from the autocovariances of squared residuals. Such an approach requires these squared residuals to be covariance stationary. Lemma 2 provides necessary and sufficient conditions for this result and is closely related to Theorem 1 of Hafner (2003). (12) shows that the autocovariances of $\{Y_t^2\}$ identify the ARCH(1) but not the GARCH(1,1) model. With respect to the latter, these autocovariances do compliment identification of β_0 conditional on the results from Lemma 1.

The moment conditions in (10)–(12) imply that the standard GMM estimator of Hansen (1982) can be used to obtain $\widehat{\theta}$. For the observed data $\{Y_t,\ t=1,\ldots,T\}$, let $X_{t-2} = \left[Y_{t-2}\cdots Y_{t-K}\right]'$ and $Z_{t-2} = \left[Y_{t-2}^2-\sigma^2\cdots Y_{t-K}^2-\sigma^2\right]'$ for $k \geq 2$. Consider the vector functions

$$g_1\left(Y_1,\ldots,Y_T;\ \theta\right) = Y_t^2-\sigma^2 \tag{13}$$

$$g_2\left(Y_1,\ldots,Y_T;\ \theta\right) = \left(Y_t^2-\sigma^2\right)Y_{t-1}-\alpha Y_t^3$$

$$g_3\left(Y_1,\ldots,Y_T;\ \theta\right) = \left(Y_t^2-\sigma^2\right)\left(X_{t-2}-(\alpha+\beta)X_{t-1}\right)$$

$$g_4\left(Y_1,\ldots,Y_T;\ \theta\right) = \left(Y_t^2-\sigma^2\right)\left(Z_{t-2}-(\alpha+\beta)Z_{t-1}\right)$$

and stack them into a single vector $g\left(\cdot;\ \theta\right)$. An estimator for θ can then be defined as

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\left[T^{-1}\sum_{t=1}^{T}g\left(\cdot;\ \theta\right)\right]' W_T\left[T^{-1}\sum_{t=1}^{T}g\left(\cdot;\ \theta\right)\right], \tag{14}$$

for some sequence of positive definite W_T. The sample moments $T^{-1}\sum_{t=1}^{T}g_2\left(\cdot;\ \theta\right)$ and $T^{-1}\sum_{t=1}^{T}g_3\left(\cdot;\ \theta\right)$ reflect the restrictions imposed by the conditional variance model in (3) on the degree of asymmetry in $\{Y_t\}$. Similarly, the sample moments $T^{-1}\sum_{t=1}^{T}g_4\left(\cdot;\ \theta\right)$ summarize the restrictions of (3) on the autocovariances of $\{Y_t^2\}$ that, of course, imply restrictions on the fourth moment of $\{Y_t\}$.

By utilizing information from the third and fourth moments, (14) relates to the Quadratic M-Estimators of Meddahi and Renault (1997) and the efficient GMM estimator of Skoglund (2001).[3] Given Theorem 4.2 in Meddahi and Renault, (14) can even be efficient conditional on a given filtration of the information set available at $t-1$. For instance, if I_{t-1} is the set of information available at $t-1$, and $J_{t-1} \subset I_{t-1}$, then if J_{t-1} preserves the parametric form of (3) and renders $E\left[Y_t^i \mid J_{t-1}\right]$ constant for $i = 3, 4$, then (14) would be efficient with respect to the third and fourth moments. In general, however, the use of the third and fourth moments in (14) will tend to correspond with some loss of efficiency because these moments will tend to vary with respect to I_{t-1}. This loss of efficiency is less of a concern in this paper as is the construction of simple estimators for the GARCH(1,1) model, and the inclusion of conditional third and fourth moments greatly complicates any GMM estimator.

3. A Theorem and Implications

Substitution of (6) into (5) yields

$$\widetilde{Y}_t^2 = \alpha_0 \widetilde{Y}_{t-1}^2 + R_t; \quad R_t = W_t + \alpha_0 \sum_{i=1}^{t-1} \beta_0^i \widetilde{Y}_{t-1-i}^2 + \beta_0^t \widetilde{h}_0, \tag{15}$$

a result that is useful for establishing a sequence of linear estimators for θ (See Corollary 1).

THEOREM. *For the model of (2) and (3), consider the estimator in (14). Let Assumptions A1–A3 hold, and assume that $W_T \xrightarrow{p} W_0$, a positive definite matrix. Then $\widehat{\theta} \xrightarrow{p} \theta_0$.*

The Theorem establishes a weakly consistent GMM estimator of the univariate GARCH(1,1) model that is based on the asymmetry of $\{Y_t\}$ and the autocovariances of $\{Y_t^2\}$. If $W_T = W_T\left(\widetilde{\theta}\right)$, where $\widetilde{\theta}$ is some preliminary consistent estimate of θ_0, then (14) is the familiar two-step GMM estimator. If $W_T = W_T(\theta)$, then (14) is the CUE of Hansen, Heaton, and Yaron (1996). Depending on the choice of K, the number of moment conditions in (14) can be large. While the use of many moment conditions leads to higher asymptotic efficiency, it can also lead to higher bias in the two-step GMM estimator (see, e.g., Han and Phillips 2005). The CUE has a relatively smaller bias (see Newey and Windmeijer 2005).

[3] First discussed by Hansen (1982), efficient GMM estimation utilizes the optimal choice of instruments from a set of conditional moment restrictions.

The Theorem assumes a stationary fourth moment. Works by Weis (1986), Rich, Raymond, and Butler (1991), and Guo and Phillips (2001) all require fourth moment stationarity for consistency. As is evident from the proof of the Theorem, consistency still follows if only Assumptions A1-A2 hold and if $g(\cdot;\ \theta)$ is redefined to only include the vector functions $g_1(\cdot;\ \theta) - g_3(\cdot;\ \theta)$. In this case, third moment stationarity of $\{Y_t\}$ is a necessary condition for both identification and an application of the LLN. In the event that $\{Y_t\}$ is fourth moment stationary, (14) defines a strictly more efficient estimator than one that omits $g_4(\cdot;\ \theta)$ from $g(\cdot;\ \theta)$. However, the Theorem can still apply in cases where this fourth moment condition appears violated (see Bollerslev 1986 and Zadrozny 2005).

Let a_i be the element from the ith row of a row vector a, and A_{ij} be the element from the ith row and jth column of a matrix A. Adapting the efficient GMM estimator of Skoglund (2001) to the model of (2) and (3) produces

$$\widehat{\vartheta} = \underset{\vartheta \in \Theta}{\arg\min} \left[T^{-1} \sum_{t=1}^{T} f(\cdot;\ \vartheta) \right]' \Lambda_T(\vartheta) \left[T^{-1} \sum_{t=1}^{T} f(\cdot;\ \vartheta) \right],$$

where

$$f_i(\cdot;\ \vartheta) = \frac{1}{\Delta_t} \left(\frac{\partial h_t}{\partial \vartheta_i} \right) h_t^{-1} \left[\left(\frac{Y_t}{h_t^{1/2}} \right) E\left[Y_t^3 \mid \Psi_{t-1} \right] - \left(\left(\frac{Y_t^2}{h_t} \right) - 1 \right) \right],$$

$$\Delta_t = \left(E\left[Y_t^4 \mid \Psi_{t-1} \right] - 1 \right) - E\left[Y_t^3 \mid \Psi_{t-1} \right]^2,$$

and $\Lambda_T(\vartheta) = \left(T^{-1} \sum_{t=1}^{T} f(\cdot;\ \vartheta) f(\cdot;\ \vartheta)' \right)^{-1}$ for $i = 1, 2, 3$. The moments from $\widehat{\vartheta}$ depend on both the third and fourth moment of $\{Y_t\}$ conditional on Ψ_{t-1} as well as on derivatives of the conditional variance function. In contrast, the moments from (14), while implied by the conditional variance function, do not take this function as an explicit input. In addition, these moments depend on the third and fourth moments of $\{Y_t\}$ only unconditionally. Therefore, while less efficient than $\widehat{\vartheta}$, $\widehat{\theta}$ is much simpler to implement. The following two corollaries further bolster this claim by showing that estimation of $\widehat{\theta}$ is possible through a sequence of linear estimators.

COROLLARY 1. *Consider* $\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} Y_t^2$. *Let* $\overline{g}\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right) = \begin{bmatrix} g_3\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right) \\ g_4\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right) \end{bmatrix}$, *where* $g_3\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)$ *and* $g_4\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)$ *are defined in (13). Let Assumptions A1–A3* *hold for the model of (2) and (3). Consider*

$$\widehat{\alpha} = \left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t-1}^{2} Y_{t-1} \right)^{-1} \sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t}^{2} Y_{t-1}, \tag{16}$$

where $\widehat{\widetilde{Y}}_{t}^{2} = Y_t^2 - \widehat{\sigma}^2$, *and*

$$\widehat{\beta} = \arg\min_{\beta \in \Theta} \left[T^{-1} \sum_{t=1}^{T} \overline{g} \left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta \right) \right]' \overline{W}_T \left[T^{-1} \sum_{t=1}^{T} \overline{g} \left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta \right) \right] \tag{17}$$

for some sequence of positive definite $\overline{W}_T$. *Assume that* $\overline{W}_T \xrightarrow{p} \overline{W}_0$, *a positive definite matrix. Then* $\widehat{\alpha} \xrightarrow{p} \alpha_0$ *and* $\widehat{\beta} \xrightarrow{p} \beta_0$. *Furthermore, if* $\overline{W}_T = \overline{W}_T \left(\widetilde{\beta} \right)$, *where* $\widetilde{\beta}$ *is a consistent preliminary estimate of* β_0, *then*

$$\widehat{\beta} = \left(\left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t}^{2} \widehat{U}_{t-1} \right)' \overline{W}_T \left(\widetilde{\beta} \right) \left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t}^{2} \widehat{U}_{t-1} \right) \right)^{-1} \left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t}^{2} \widehat{U}_{t-1} \right)' \overline{W}_T \left(\widetilde{\beta} \right) \left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t}^{2} \widehat{U}_{t-2} \right) - \widehat{\alpha}, \tag{18}$$

where $\widehat{U}_{t-2} = \begin{pmatrix} X_{t-2} \\ \widehat{Z}_{t-2} \end{pmatrix}$.

The power of Corollary 1 is the realization that estimation of α_0 and β_0 can be conducted separately and that this separation affords a linear estimator for each. (16) is the feasible linear TSLS estimator of α_0 in (15), where Y_{t-1} serves as the instrument for $\widetilde{Y}_{t-1}^2$. (18) is the solution to the two-step GMM estimator in (17), also linear. Calculating $\widehat{\sigma}^2$ first, then $\widehat{\alpha}$ by (16) and, finally, $\widehat{\beta}$ by (18), permits $\widehat{\theta}$ to be obtained without the need for numerical optimization techniques and consistent starting values. If $\overline{W}_T = \overline{W}_T(\beta)$, then (17) is no longer linear. However, $\widehat{\beta}$ can still be easily obtained via a grid search, thereby avoiding the need to calculate numerical derivatives and the potential problem of finding local minima.

COROLLARY 2 *Consider* $\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} Y_t^2$. *Let Assumptions A1–A3 hold for the model of (2) and (3), and assume that* $\beta_0 = 0$. *Consider*

$$\widehat{\alpha} = \arg\min_{\alpha \in \Theta} \left[T^{-1} \sum_{t=1}^{T} \left(\widehat{\widetilde{Y}}_{t}^{2} - \alpha\, \widehat{\widetilde{Y}}_{t-1}^{2} \right) \widehat{U}_{t-1} \right]' \Omega_T \left[T^{-1} \sum_{t=1}^{T} \left(\widehat{\widetilde{Y}}_{t}^{2} - \alpha\, \widehat{\widetilde{Y}}_{t-1}^{2} \right) \widehat{U}_{t-1} \right], \tag{19}$$

where $\widehat{\widetilde{Y}}_t^2 = Y_t^2 - \widehat{\sigma}^2$ *and* $\widehat{U}_{t-1} = \begin{pmatrix} X_{t-1} \\ \widehat{Z}_{t-1} \end{pmatrix}$ *for some sequence of positive definite* Ω_T*. Assume that* $\Omega_T \xrightarrow{p} \Omega_0$*, a positive definite matrix. Then* $\widehat{\alpha} \xrightarrow{p} \alpha_0$*. Furthermore, if* $\Omega_T = \Omega_T(\widetilde{\alpha})$*, where* $\widetilde{\alpha}$ *is some consistent preliminary estimate of* α*, then*

$$\widehat{\alpha} = \left(\left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t-1}^2 \widehat{U}_{t-1} \right)' \Omega_T(\widetilde{\alpha}) \left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t-1}^2 \widehat{U}_{t-1} \right) \right)^{-1} \left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t-1}^2 \widehat{U}_{t-1} \right)' \Omega_T(\widetilde{\alpha}) \left(\sum_{t=1}^{T} \widehat{\widetilde{Y}}_{t}^2 \widehat{U}_{t-1} \right). \tag{20}$$

If $\Omega_T = I$, then Corollary 2 supports TSLS estimation of (15) using U_{t-1} as instruments for $\widetilde{Y}_{t-1}^2$. (20) nests the OLS estimator of Weis (1986) and the IV estimator of Rich, Raymond, and Butler (1991) where lags of the squared residuals comprise the instrument vector.[4] (20) should be strictly more efficient than either of these, however, owing to the consideration of the third moment. (20) is also more general since it does not require fourth moment stationarity for consistency. If $\Omega_T = \Omega_T(\alpha)$, then (19) links univariate ARCH estimation to the class of GEL estimators introduced by Smith (1997).

From Hansen (1982), the optimal GMM weighting matrix is the inverse of the variance-covariance matrix of the moment conditions. In the context of (14), (17), or (19), however, consistency of this optimal weighting matrix requires $\{Y_t\}$ to be eighth moment stationary. For many applications in financial economics, this assumption proves overly restrictive. Of course, the identity matrix supports consistency of the proposed estimators. A question is, therefore, to what extent can a data dependent weighting matrix improve finite sample efficiency?

For the moment conditions $E[g(\cdot;\, \theta_0)]$ where $g(\cdot;\, \theta_0) = (g_i(\cdot;\, \theta_0))$ for $i = 1, \ldots, 2K$, the optimal weighting matrix is $E\left[g(\cdot;\, \theta_0)\, g(\cdot;\, \theta_0)'\right]^{-1}$, assuming that $\{g(\cdot;\, \theta_0)\}$ is not autocorrelated. Preventing the use of this weighting matrix is a concern over the existence of moments. A natural choice for an alternative weighting matrix would involve a robust analog to $E\left[g(\cdot;\, \theta_0)\, g(\cdot;\, \theta_0)'\right]$. Towards that end, consider the matrix $W_\rho(\theta_0) = \left(w_{ij,\rho}(\theta_0)\right)$, where $w_{ij,\rho}(\theta_0)$ is Spearman's (1904) rho-statistic measured between $g_i(\cdot;\, \theta_0)$ and $g_j(\cdot;\, \theta_0)$. Alterna-

[4] Corollary 2 is stated in terms of the ARCH(1) model. Extension to the ARCH(p) case, however, is completely straightforward. Specification of the semi-strong GARCH model in (2) and (3) does not reflect this fact because the focus of this paper is on standard GMM estimation of univariate GARCH models, and Theorem 1 does not extend to GARCH(p,q) models where p,q $\geq$ 1. For a general GARCH(p,q) model, the presence of skewness is not sufficient for GMM identification. Causing this insufficiency is a lack of suitable instruments.

tively, one can consider the matrix $W_{\tau}(\theta_0) = (w_{ij,\tau}(\theta_0))$, where $w_{ij,\tau}(\theta_0)$ is Kendall's (1938) tau-statistic measured between the same moment conditions. Each of these two statistics is a rank dependent measure of correspondence ranging between -1 and 1. Therefore, $W_{\rho}(\theta_0)$ or $W_{\tau}(\theta_0)$ is a robust correlation matrix since according to Taskinen, Oja, and Randles (2005), even assumptions regarding the existence of the first moments of $g(\cdot;\ \theta_0)$ are not needed for consistency of either statistic.

Similar to Weiss (1986) and Rich, Raymond, and Butler (1991), the estimators in the theorem and corollaries can be shown to be asymptotically normal if $\{Y_t\}$ is eighth moment stationary. Given the restrictive nature of this assumption, standard errors for $\widehat{\theta}$ can alternatively be generated by the parametric bootstrap. Suppose that the data generating process for $\{Y_t\}$ is characterized by (2), (3), and (9) where $E\left[\epsilon_t \mid \Psi_{t-1}\right] = 0$ and $E\left[\epsilon_t^2 \mid \Psi_{t-1}\right] = 1$, which is the semi-strong GARCH model of Lee and Hansen (1994) and Escanciano (2009). Using one of the estimators described above, obtain $\widehat{h}_t$. Then $\widehat{\epsilon}_t = Y_t/\sqrt{\widehat{h}_t}$. Apply the nonoverlapping block bootstrap method of Carlstein (1986) to these standardized residuals to obtain the bootstrap sample $\widehat{\epsilon}_t^*$. Use these bootstrap residuals to construct the series $\widehat{Y}_t^* = \sqrt{\widehat{h}_t^*}\widehat{\epsilon}_t^*$, where $\widehat{h}_t^*$ depends on the parameter estimates from the original data sample. Estimate the model of (2) and (3) on $\widehat{Y}_t^*$, making sure to center the bootstrap moment conditions with the original parameter estimates as in Hall and Horowitz (1996). Repetition of this procedure permits the calculation of bootstrap standard errors for $\widehat{\theta}$ that are robust to higher moment dynamics in ϵ_t.[5] This same procedure can also be used to bootstrap the GMM objective function as discussed in Brown and Newey (2002) for a non-parametric test of overidentifying restrictions that speaks to the fit of the GARCH(1,1) model to the given data under study.

4. Monte Carlo

Consider the data generating process in (2), (3), and (9), where ϵ_t is a standardized Gamma(2,1) random variable. This DGP is one of strong GARCH. The skewness and kurtosis of ϵ_t is $2/\sqrt{2}$ and 6, respectively. All simulations are conducted across 1,000 trials with sample sizes ranging from 5,000 to 40,000 observations. In each simulation, the first 200 observations are dropped in

[5]Escanciano (2009) shows that fourth moment dependence of ϵ_t impacts the calculation of standard errors for the QMLE.

order to avoid initialization effects. Because of a concern over the existence of moments, summary statistics for the parameter estimates are robust measures of bias and dispersion. The standard deviation of the parameter estimates is also reported which, while not a robust measure, gives an indication of the effects of outliers.

Table 1 summarizes the results from simulations of a GARCH(1,1) model. The values for α_0 and β_0 are chosen to reflect the low ARCH and high GARCH terms frequently encountered in empirical studies. Five different estimators are considered: (1) the QMLE of $\widehat{\vartheta}$; (2) the CUE of $\widehat{\theta}$; (3) the traditional two-step GMM estimator of $\widehat{\theta}$ (GMM); (3) the multi-step estimator of $\widehat{\sigma}^2$ by OLS, $\widehat{\alpha}$ by TSLS, and $\widehat{\beta}$ by CUE (OLS/TSLS/CUE)[6]; (4) the multi-step estimator of $\widehat{\sigma}^2$ by OLS, $\widehat{\alpha}$ by TSLS, and $\widehat{\beta}$ by GMM (OLS/TSLS/GMM). The QMLE serves as a benchmark. For the CUE and GMM estimators, the weighting matrix is the robust correlation matrix formed using Spearman's-rho.[7] The applications of CUE and GMM set $K = 10$. This value was chosen because it tended to minimize the bias-variance trade-off from increasing the lag order of the GMM estimator.

A significant finding is that QMLE does not dominate the simple GMM estimators. As evidenced in Table 1, $\widehat{\sigma}$ and $\widehat{\alpha}$ have the same biases, smaller median absolute errors, and smaller decile ranges when estimated with CUE as opposed to QMLE. The dispersion of the CUE can be heightened relative to comparable estimators as seen, for example, through a comparison of both the decile ranges and standard deviations of the OLS/TSLS/CUE and OLS/TSLS/GMM estimates. This finding compliments simulation evidence presented in Hansen, Heaton, and Yaron (1996). Of the simple GMM estimators, CUE is associated with the smallest biases. This statement is most apparent for $\widehat{\beta}$, where GMM and OLS/TSLS/GMM have biases nearly twice as large as CUE and OLS/TSLS/CUE. Also apparent from $\widehat{\beta}$ is a tendency for the simple GMM estimators as a group to display higher biases than QMLE. For the GMM and OLS/TSLS/GMM estimators, these heightened biases are particularly acute. However, these biases significantly dissipate with an increasing sample size as is evidenced by the results in Table 2. Here, small and uniformly decreasing biases are shown for the OLS/TSLS/GMM estimator. Uniformly decreasing levels of dispersion in the parameter estimates are evidenced as well. Recall from Corollary 1 that OLS/TSLS/GMM utilizes

[6] $\widehat{\sigma}^2$ is obtained from a regression of Y_t^2 on a constant.

[7] Simulations (not reported here) also considered the robust correlation matrix formed with Kendall's-tau. Results for the two weighting matrices were very similar. Since Kendall's-tau is computationally expensive, Spearman's-rho is used instead.

a linear estimator at each step to obtain $\widehat{\theta}$. The results of Table 2, thus, support simple GMM estimators as advantageous alternatives for GARCH(1,1) model estimation on very high frequency data as is commonly analyzed in the market microstructure literature, where studies of intra-daily returns can involve sample sizes of nearly 100,000 observations (see, e.g., Anderson and Bollerslev 1997). At relatively lower sample sizes, the results of Table 1 support the use of CUE and OLS/TSLS/CUE over GMM and the fully linear OLS/TSLS/GMM estimator.

Table 3 summarizes the results from simulations of an ARCH(1) model. Two additional estimators are considered: (1) the two-step estimator of $\widehat{\sigma}^2$ by OLS and $\widehat{\alpha}$ by IV (OLS/IV)[8]; (2) the two-step estimator of $\widehat{\sigma}^2$ by OLS and $\widehat{\alpha}$ by OLS (OLS/OLS). This second estimator is studied by Weis (1986). Of the moment-based estimators, OLS/CUE displays the smallest bias, but it does not dominate in terms efficiency as measured by the decile range. Moreover, CUE displays the largest bias of all the estimators of $\widehat{\sigma}$. OLS/IV is marginally better than OLS/OLS in terms of bias and dispersion, but there is no noticeable efficiency gain moving from an IV estimator to a GMM estimator of $\widehat{\alpha}$. This result is odd since other simulations not reported here for the GARCH(1,1) model showed significant improvements in terms of both bias and dispersion reduction from moving to a GMM estimator with a data dependent weighting matrix from a GMM estimator with the identity matrix. Finally, for the ARCH(1) model, QMLE dominates in terms of bias and efficiency.

Tables 4 and 5 summarize the simulation results of an ARCH(1) and GARCH(1,1) model, neither of which have a finite fourth moment according to the inequality restriction of Zadrozny (2005).[9] For the GARCH(1,1) model, only the QMLE and CUE are considered. For the ARCH(1) model, OLS/OLS is also considered as a means of judging the finite sample effects of naively applying an inconsistent estimator. For the GARCH(1,1) model, QMLE once again fails to dominate. While having a higher bias, $\widehat{\sigma}$ has a lower median absolute error and decile range when estimated by the CUE. QMLE does dominate, however, and rather significantly, in estimating $\widehat{\alpha}$ and $\widehat{\beta}$. For the ARCH(1) model, the CUE dominates OLS/OLS, but QMLE dominates the CUE.

[8] IV estimation of $\widehat{\alpha}$ is equivalent to (20) with $\Omega_T = I$.

[9] Parameter values are chosen such that this inequality restriction is just violated so as to maximize the likelihood of a finite third moment.

5. Conclusion

The main contribution of this paper is to provide simple, weakly consistent, GMM estimators for the GARCH(1,1) model. These estimators rely on unconditional skewness but do not require treatment of the third and fourth conditional moments. Moreover, these estimators require less strict moment existence assumptions then ARCH estimators based upon the autocovariances of squared residuals. Linear versions of these estimators facilitate GARCH(1,1) estimation on very high frequency data and on moderately sized (in the time dimension) data sets where many such models need to be estimated, as is common in portfolio optimization and Value at Risk (VaR) problems faced by financial industry professionals. Nonlinear versions of these estimators can outperform QMLE in finite samples. Finally, these estimators compliment conditional asset pricing tests that rely on standard GMM procedures. A question for future research is whether these simple estimators when applied to intra-day financial return data and aggregated to a lower sampling frequency (say, daily or monthly) using the results of Drost and Nijman (1993) outperform the QMLE applied at the lower frequency either in terms of bias and efficiency of the parameter estimates or in terms of out-of-sample fit of the conditional volatility forecasts.

Appendix

PROOF OF LEMMA 1: Given mean stationarity of $\{W_t Y_t\}$, and the result from (8),

$$\begin{aligned} E\left[\widetilde{Y}_t^2 Y_{t-1}\right] &= E\left[\left(\widetilde{h}_t + W_t\right) Y_{t-1}\right] \\ &= E\left[\left(\alpha_0 \widetilde{Y}_{t-1}^2 + \beta_0 \widetilde{h}_{t-1}\right) Y_{t-1}\right] \\ &= \alpha_0 E\left[W_t Y_t\right]. \end{aligned} \tag{21}$$

Since

$$\begin{aligned} E\left[\widetilde{Y}_t^2 Y_{t-2}\right] &= E\left[\widetilde{h}_t Y_{t-2}\right] \\ &= \alpha_0 E\left[\widetilde{Y}_{t-1}^2 Y_{t-2}\right] + \beta_0 E\left[\widetilde{h}_{t-1} Y_{t-2}\right] \\ &= \left(\alpha_0 + \beta_0\right) E\left[\widetilde{Y}_{t-1}^2 Y_{t-2}\right], \end{aligned}$$

and

$$E\left[\widetilde{Y}_{t-1}^2 Y_{t-2}\right] = \alpha_0 E\left[W_t Y_t\right]$$

given mean stationarity of $\{W_t Y_t\}$ again, then

$$E\left[\widetilde{Y}_t^2 Y_{t-2}\right] = \alpha_0 \left(\alpha_0 + \beta_0\right) E\left[W_t Y_t\right].$$

Repeated applications of recursive substitution into $E\left[\widetilde{Y}_t^2 Y_{t-k}\right]$ reveals that

$$E\left[\widetilde{Y}_t^2 Y_{t-k}\right] = \alpha_0 \left(\alpha_0 + \beta_0\right)^{k-1} E\left[W_t Y_t\right]. \tag{22}$$

Solving (22) for $k = k + 1$ and comparing the result to $E\left[\widetilde{Y}_t^2 Y_{t-k}\right]$ produces (11).■

PROOF OF LEMMA 2: From (5) follows that

$$E\left[\widetilde{Y}_t^4\right] = E\left[\left(\widetilde{h}_t + W_t\right)^2\right] = E\left[\widetilde{h}_t^2\right] + E\left[W_t^2\right].$$

Given (4),

$$E\left[\widetilde{h}_t^2\right] = (\alpha_0+\beta_0)^2 E\left[\widetilde{h}_{t-1}^2\right] + \alpha_0^2\lambda_0. \tag{23}$$

Recursive substitution into (23) produces

$$E\left[\widetilde{h}_t^2\right] = \left(1+(\alpha_0+\beta_0)^2+\cdots+(\alpha_0+\beta_0)^{2(\tau-1)}\right)\alpha_0^2\lambda_0 + (\alpha_0+\beta_0)^{2\tau}E\left[\widetilde{h}_{t-\tau}^2\right]$$

for $\tau \geq 1$. It is well known that $(\alpha_0+\beta_0)^{2\tau} \to 0$ as $\tau \to \infty$ if and only if $\alpha_0+\beta_0 < 1$. Therefore, $E\left[\widetilde{h}_t^2\right] \to \left(\frac{\alpha_0^2}{1-(\alpha_0+\beta_0)^2}\right)\lambda_0$ as $\tau \to \infty$ if and only if A2 holds. Let $E\left[\widetilde{h}_t^2\right] = \eta_0$. For $k=1$,

$$\begin{aligned} E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-1}^2\right] &= E\left[E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-1}^2 \mid \Psi_{t-1}\right]\right] \\ &= E\left[\left(\alpha_0\widetilde{Y}_{t-1}^2+\beta_0\widetilde{h}_{t-1}\right)\widetilde{Y}_{t-1}^2\right] \\ &= \alpha_0\lambda_0 + (\alpha_0+\beta_0)\eta_0 \end{aligned}$$

For $k \geq 2$,

$$\begin{aligned} E\left[\widetilde{h}_t \mid \Psi_{t-k}\right] &= \alpha_0 E\left[\widetilde{Y}_{t-1}^2 \mid \Psi_{t-k}\right] + \beta_0 E\left[\widetilde{h}_{t-1} \mid \Psi_{t-k}\right] \\ &= (\alpha_0+\beta_0)E\left[\widetilde{h}_{t-1} \mid \Psi_{t-k}\right] \\ &= (\alpha_0+\beta_0)^2 E\left[\widetilde{h}_{t-2} \mid \Psi_{t-k}\right] \\ &\vdots \\ &= (\alpha_0+\beta_0)^{\tau-1}E\left[h_{t-(k-1)} \mid \Psi_{t-k}\right] \\ &= (\alpha_0+\beta_0)^{\tau-1}\left[\alpha_0 Y_{t-k}^2 + \beta_0 h_{t-k}\right] \end{aligned}$$

and, therefore,

$$\begin{aligned} E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2\right] &= E\left[E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2 \mid \Psi_{t-k}\right]\right] \\ &= E\left[E\left[\widetilde{h}_t \mid \Psi_{t-k}\right]\widetilde{Y}_{t-k}^2\right] \\ &= (\alpha_0+\beta_0)^{k-1}\left[\alpha_0\lambda_0 + (\alpha_0+\beta_0)\eta_0\right]. \end{aligned} \tag{24}$$

Given (24), $E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2\right] \to 0$ as $k \to \infty$. Solving (24) for $k = k+1$ and comparing the result to $E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2\right]$ grants (12).■

LEMMA 3: Given Assumptions A1–A3, the following conditions hold:

CONDITION C1: $T^{-1}\sum_{t=1}^{T} Y_t \xrightarrow{p} 0$

CONDITION C2: $T^{-1}\sum_{t=1}^{T} Y_t^2 \xrightarrow{p} \sigma^2$

CONDITION C3: $T^{-1}\sum_{t=1}^{T} W_t \xrightarrow{p} 0$

CONDITION C4: $T^{-1}\sum_{t=1}^{T} W_t Y_t \xrightarrow{p} \gamma_0$

CONDITION C5: $T^{-1}\sum_{t=1}^{T} W_{t-l} Y_{t-k} \xrightarrow{p} 0 \;\forall\; k \neq l$

CONDITION C6: $T^{-1}\sum_{t=1}^{T} W_t W_{t-k} \xrightarrow{p} 0 \;\forall\; k \geq 1$

CONDITION C7: $T^{-1}\sum_{t=1}^{T} W_t^2 \xrightarrow{p} \lambda_0$

CONDITION C8: *For a constant C where $0 < C < 1$ and a martingale difference sequence $\{Z_t\}$ that is uniformly integrable, $T^{-1}\sum_{t=1}^{T} C^t Z_t \xrightarrow{p} 0$.*

PROOF. Given A1, Y_t is covariance stationary. C1 then follows by (2) and the LLN. Given Lemma 2, Y_t^2 is covariance stationary with $E\left[\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2\right] \to 0$ as $k \to \infty$ (see (24)). C2 then also follows from the LLN. $E\left[W_t \mid \Psi_{t-1}\right] = 0$ by construction. As a consequence, $E\left[W_t W_{t-k}\right] = 0 \;\forall\; k \geq 1$. Given A3(i), W_t is covariance stationary, and C3 follows from the LLN. Given A2(i)-(ii), C4 follows from Theorem 1 of Andrews (1988). $\left\{W_{t-l}Y_{t-k}\right\}$ and $\left\{W_t W_{t-k}\right\}$ are both martingale difference sequences. Given A2(iii) and A3(ii), Theorem 1 of Andrews (1988) applies to each to establish C5 and C6, respectively. A3(i) and A3(iii) allow C7 to follow from Theorem 1 of Andrews (1988). Lastly, since $\{Z_t\}$ is uniformly integrable, $\exists$ a $c > 0$ for every $\epsilon > 0$ such that

$$E\left[|Z_t| \times I\left(|Z_t| \geq c\right)\right] < \epsilon,$$

where $I(|Z_t| \geq c) = 1$ if $|Z_t| \geq c$ and 0 otherwise. Let $X_t = C^t Z_t$. Then

$$|X_t| = \left|C^t\right| |Z_t| < |Z_t|,$$

and

$$|X_t| \times I(|X_t| \geq c) \leq |Z_t| \times I(|Z_t| \geq c).$$

As a consequence,

$$E[|X_t| \times I(|X_t| \geq c)] < \epsilon,$$

and $\{X_t\}$ is uniformly integrable. Theorem 1 of Andrews (1988) then establishes C8. ■

PROOF OF THE THEOREM: By C2,

$$\begin{aligned} \mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T} g_1(\cdot;\ \theta)\right) &= \sigma_0^2 - \sigma^2 \\ &= E[g_1(\cdot;\ \theta)] \end{aligned} \tag{25}$$

Next,

$$\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T} g_2(\cdot;\ \theta)\right) = \mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T} Y_t^2 Y_{t-1}\right) - \alpha \mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T} Y_t^3\right)$$

by C1. Given (7),

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} Y_t^2 Y_{t-1} &= T^{-1}\sum_{t=1}^{T}\left(W_t + \alpha_0 \sum_{i=1}^{t-1}(\alpha_0+\beta_0)^{i-1} W_{t-i} + \beta_0(\alpha_0+\beta_0)^{t-1}\widetilde{h}_0 + \sigma_0^2\right) Y_{t-1} \\ &= \alpha_0 T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}(\alpha_0+\beta_0)^{i-1} W_{t-i} Y_{t-1} + \text{(3 additional terms)} \end{aligned}$$

where the probability limit for each of these additional terms is zero given C1, C5, and C8.

The term $T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}(\alpha_0+\beta_0)^{i-1}W_{t-i}Y_{t-1} =$

$$\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\left(W_{t-1}+(\alpha_0+\beta_0)W_{t-2}+(\alpha_0+\beta_0)^2W_{t-3}+\cdots+(\alpha_0+\beta_0)^{t-2}W_1\right)Y_{t-1} \\
= \; & T^{-1}\sum_{t=1}^{T}W_{t-1}Y_{t-1}+(\alpha_0+\beta_0)T^{-1}\sum_{t=1}^{T}W_{t-2}Y_{t-1}+(\alpha_0+\beta_0)^2T^{-1}\sum_{t=1}^{T}W_{t-3}Y_{t-1}+\cdots \\
& +W_1T^{-1}\sum_{t=1}^{T}(\alpha_0+\beta_0)^{t-2}Y_{t-1}
\end{aligned}$$

By C4, C5, and C8, therefore, $\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}Y_t^2Y_{t-1}\right)=\alpha_0\gamma_0$. Furthermore, since $T^{-1}\sum_{t=1}^{T}Y_t^3 = T^{-1}\sum_{t=1}^{T}Y_t^2Y_t$, it follows that

$$\begin{aligned}
\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}g_2(\cdot;\ \theta)\right) &= (\alpha_0-\alpha)\gamma_0 \\
&= E\left[g_2(\cdot;\ \theta)\right]
\end{aligned} \tag{26}$$

Define the k^{th} element of the vector $g_3(\cdot;\ \theta)$ as

$$g_{3,k}(\cdot;\ \theta)=\left(Y_t^2-\sigma^2\right)\left(Y_{t-(k+1)}-(\alpha+\beta)Y_{t-k}\right).$$

Then,

$$\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}g_3(\cdot;\ \theta)\right)=\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}Y_t^2Y_{t-(k+1)}\right)-(\alpha+\beta)\,\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}Y_t^2Y_{t-k}\right)$$

by C1. Given (7),

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}Y_t^2Y_{t-(k+1)} &= \alpha_0T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}(\alpha_0+\beta_0)^{i-1}W_{t-i}Y_{t-(k+1)}+(\text{3 additional terms}) \\
&= \alpha_0(\alpha_0+\beta_0)^kT^{-1}\sum_{t=1}^{T}W_{t-(k+1)}Y_{t-(k+1)} \\
&\quad +\alpha_0T^{-1}\sum_{t=1}^{T}\sum_{i\neq k+1}(\alpha_0+\beta_0)^{i-1}W_{t-i}Y_{t-(k+1)}+(\text{3 additional terms})
\end{aligned}$$

The three additional terms each have probability limits equal to zero given C1, C5, and C8.

Therefore, $\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}Y_t^2Y_{t-(k+1)}\right)=\alpha_0\left(\alpha_0+\beta_0\right)^k\gamma_0$, and

$$\begin{aligned}\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}g_{3,k}\left(\cdot;\ \theta\right)\right) &= \alpha_0\left[\left(\alpha_0+\beta_0\right)-\left(\alpha+\beta\right)\right]\left(\alpha_0+\beta_0\right)^{k-1}\gamma_0 \qquad (27)\\ &= E\left[g_{3,k}\left(\cdot;\ \theta\right)\right]\end{aligned}$$

Next, define the k^{th} element the vector $g_4\left(\cdot;\ \theta\right)$ as

$$g_{4,k}\left(\cdot;\ \theta\right)=\left(Y_t^2-\sigma^2\right)\left(Y_{t-(k+1)}-\sigma^2\right)-\left(\alpha+\beta\right)\left(Y_t^2-\sigma^2\right)\left(Y_{t-k}-\sigma^2\right),$$

and consider the $\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}g_{4,k}\left(\cdot;\ \theta\right)\right)$. Again relying on the interpretation of Y_t^2 as a weighted sum of current and past innovations in (7),

$$\begin{aligned}T^{-1}\sum_{t=1}^{T}Y_t^2Y_{t-k}^2 &= \left(\sigma_0^2\right)^2+\alpha_0T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}W_{t-k}\\ &\quad+\alpha_0^2T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{t-1}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}\right)\left(\sum_{j=1}^{t-(k+1)}\left(\alpha_0+\beta_0\right)^{j-1}W_{t-k-j}\right)\\ &\quad+\left(\text{6 additional terms}\right)\\ &= \left(\sigma_0^2\right)^2+\alpha_0T^{-1}\left[\left(\alpha_0+\beta_0\right)^{k-1}\sum_{t=1}^{T}W_{t-k}^2+\sum_{t=1}^{T}\sum_{i\neq k}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}W_{t-k}\right]\\ &\quad+\alpha_0^2T^{-1}\left[\sum_{t=1}^{T}\sum_{i\neq j}\left(\alpha_0+\beta_0\right)^{(i+j)-2}W_{t-i}W_{t-k-j}+\sum_{t=1}^{T}\sum_{j=k}^{t-1}\left(\alpha_0+\beta_0\right)^{2j-k}W_{t-j-1}^2\right]\\ &\quad+\left(\text{6 additional terms}\right)\end{aligned}$$

C3, C6, and C8 are used to show that the probability limits of the 6 additional terms are each zero. $\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}W_{t-k}^2\right)=\lambda_0$ given C7.

$$\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\sum_{i\neq k}\left(\alpha_0+\beta_0\right)^{i-1}W_{t-i}W_{t-k}\right)=\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\sum_{i\neq j}\left(\alpha_0+\beta_0\right)^{(i+j)-2}W_{t-i}W_{t-k-j}\right)=0$$

given C6. The term $T^{-1}\sum_{t=1}^{T}\sum_{j=k}^{t-1}(\alpha_0+\beta_0)^{2j-k}W_{t-j-1}^2 =$

$$
\begin{aligned}
& T^{-1}\sum_{t=1}^{T}\left((\alpha_0+\beta_0)^{k}W_{t-k-1}^2+(\alpha_0+\beta_0)^{k+2}W_{t-k-2}^2+\cdots+(\alpha_0+\beta_0)^{2t-(k+2))}W_1^2\right)\\
= \; & (\alpha_0+\beta_0)^{k}T^{-1}\sum_{t=1}^{T}W_{t-k-1}^2+(\alpha_0+\beta_0)^{k+2}T^{-1}\sum_{t=1}^{T}W_{t-k-2}^2+\cdots+(\alpha_0+\beta_0)^{2t-(k+2)}W_1^2
\end{aligned}
$$

By C7, $\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\sum_{j=k}^{t-1}(\alpha_0+\beta_0)^{2j-k}W_{t-j-1}^2\right) =$

$$
\begin{aligned}
& (\alpha_0+\beta_0)^{k}\lambda_0\left(1+(\alpha_0+\beta_0)^2+(\alpha_0+\beta_0)^4+\cdots\right)\\
= \; & (\alpha_0+\beta_0)^{k}\frac{\lambda_0}{1-(\alpha_0+\beta_0)^2}
\end{aligned}
$$

and

$$
\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}Y_t^2Y_{t-k}^2\right)=\left(\sigma_0^2\right)^2+(\alpha_0+\beta_0)^{k-1}\left(\alpha_0\lambda_0+(\alpha_0+\beta_0)\eta_0\right),
$$

where $\eta_0 = E\left[\widetilde{h}_t^2\right]$ from Lemma 2. Therefore,

$$
\begin{aligned}
\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}g_{4,k}(\cdot;\,\theta)\right) &= \left(\sigma_0^2-\sigma^2\right)^2(1-(\alpha+\beta))+ \\
&\quad ((\alpha_0+\beta_0)-(\alpha+\beta))(\alpha_0+\beta_0)^{k-1}\left(\alpha_0\lambda_0+(\alpha_0+\beta_0)\eta_0\right)\\
&= E\left[g_{4,k}(\cdot;\,\theta)\right]
\end{aligned}
\tag{28}
$$

Given (25)–(28), $T^{-1}\sum_{t=1}^{T}g(\cdot;\,\theta)\xrightarrow{p}E[g(\cdot;\,\theta)]$. Let $Q(\cdot;\,\theta)=E[g(\cdot;\,\theta)]'\,W_0E[g(\cdot;\,\theta)]$, and $\widehat{Q}_T(\cdot;\,\theta)=\widehat{g}_T(\cdot;\,\theta)'\,W_T\widehat{g}_T(\cdot;\,\theta)$, where $\widehat{g}_T(\cdot;\,\theta)=T^{-1}\sum_{t=1}^{T}g(\cdot;\,\theta)$. Then $\widehat{Q}_T(\cdot;\,\theta)\xrightarrow{p}Q(\cdot;\,\theta)$ by continuity of multiplication. From (25), $E[g_1(\cdot;\,\theta)]=0$ if and only if $\sigma^2=\sigma_0^2$. From (26), $E[g_2(\cdot;\,\theta)]=0$ if and only if $\alpha=\alpha_0$ since $\gamma_0\neq 0$. If $\sigma^2=\sigma_0^2$ and $\alpha=\alpha_0$, then $E[g_3(\cdot;\,\theta)]=0$ if and only if $\beta=\beta_0$ given (27) and the fact that $\alpha_0+\beta_0$ is strictly positive. Similarly, $E[g_4(\cdot;\,\theta)]=0$ if and only if $\beta=\beta_0$ given (28) and the fact that $\alpha_0\lambda_0+(\alpha_0+\beta_0)\eta_0$ is strictly positive. Therefore, the only $\theta\in\Theta$ that satisfies $E[g(\cdot;\,\theta)]=0$ is $\theta=\theta_0$ and, as a consequence, $Q(\cdot;\,\theta)$ is uniquely minimized at $\theta=\theta_0$.■

PROOF OF COROLLARY 1: Given (15),

$$\widehat{\widetilde{Y}}_t^2 = \alpha_0 \widehat{\widetilde{Y}}_{t-1}^2 + \overline{R}_t; \quad \overline{R}_t = (\alpha_0 - 1)\left(\widehat{\sigma}^2 - \sigma_0^2\right) + R_t. \tag{29}$$

Substitution of (29) into (16) produces

$$\widehat{\alpha} = \alpha_0 + \left(T^{-1}\sum_{t=1}^{T}\widehat{\widetilde{Y}}_{t-1}^2 Y_{t-1}\right)^{-1}\left((\alpha_0 - 1)\left(\widehat{\sigma}^2 - \sigma_0^2\right)T^{-1}\sum_{t=1}^{T}Y_{t-1} + T^{-1}\sum_{t=1}^{T}R_t Y_{t-1}\right).$$

$$\begin{aligned}
\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\widehat{\widetilde{Y}}_{t-1}^2 Y_{t-1}\right) &= \operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}Y_{t-1}^3\right) + \operatorname{p\,lim}\left(\widehat{\sigma}^2\right)\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}Y_{t-1}\right) \\
&= \gamma_0
\end{aligned}$$

given C1, C2, and (26) in the proof of the Theorem. As a result,

$$\operatorname{p\,lim}\widehat{\alpha} = \alpha_0 + \gamma_0^{-1}\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}R_t Y_{t-1}\right).$$

Given the definition of R_t in (15),

$$T^{-1}\sum_{t=1}^{T}R_t Y_{t-1} = T^{-1}\sum_{t=1}^{T}W_t Y_{t-1} + T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\beta_0^i \widetilde{Y}_{t-1-i}^2 Y_{t-1} + \widetilde{h}_0 T^{-1}\sum_{t=1}^{T}\beta_0^t Y_{t-1}.$$

The first and third terms in this expression converge weakly towards zero given C5 and C8, respectively. From (7),

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\beta_0^i \widetilde{Y}_{t-1-i}^2 Y_{t-1} &= \widetilde{h}_0 T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\beta_0^i\left(\alpha_0 + \beta_0\right)^{t-2-i}Y_{t-1} + T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}W_{t-1-i}Y_{t-1} \\
&\quad + \alpha_0 T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-1}\sum_{j=1}^{t-2-i}\left(\alpha_0 + \beta_0\right)^{j-1}W_{t-1-i-j}Y_{t-1}
\end{aligned}$$

Applications of C5 and C8 again establishes $\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}R_t Y_{t-1}\right) = 0$, from which the result $\widehat{\alpha} \xrightarrow{p} \alpha_0$ then follows.

Next, let

$$g_3\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right) = \left(Y_t^2 - \widehat{\sigma}^2\right)\left(X_{t-2} - \left(\widehat{\alpha}+\beta\right) X_{t-2}\right),$$

where the k^{th} element of this vector is defined as

$$g_{3,k}\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right) = \left(Y_t^2 - \widehat{\sigma}^2\right)\left(Y_{t-(k+1)} - \left(\widehat{\alpha}+\beta\right) Y_{t-k}\right).$$

Since $\widehat{\alpha} \xrightarrow{p} \alpha_0$ and given C1 and C2,

$$\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} g_{3,k}\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)\right) = \operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} Y_t^2 Y_{t-(k+1)}\right) - \left(\alpha_0+\beta\right)\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} Y_t^2 Y_{t-k}\right).$$

Furthermore, since $\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} Y_t^2 Y_{t-(k+1)}\right) = \alpha_0\left(\alpha_0+\beta_0\right)^k \gamma_0$ as demonstrated in the proof of the Theorem,

$$\begin{aligned}\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} g_{3,k}\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)\right) &= \alpha_0\left(\beta_0-\beta\right)\left(\alpha_0+\beta_0\right)^{k-1}\gamma_0 \\ &= E\left[g_{3,k}\left(\cdot;\ \sigma_0^2,\ \alpha_0,\ \beta\right)\right]\end{aligned} \tag{30}$$

Let

$$g_4\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right) = \left(Y_t^2 - \widehat{\sigma}^2\right)\left(\widehat{Z}_{t-2} - \left(\widehat{\alpha}+\beta\right)\widehat{Z}_{t-2}\right),$$

where the k^{th} element of this vector is defined as

$$g_{4,k}\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right) = \left(Y_t^2 - \widehat{\sigma}^2\right)\left(\left(Y_{t-(k+1)}^2 - \widehat{\sigma}^2\right) - \left(\widehat{\alpha}+\beta\right)\left(Y_{t-k}^2 - \widehat{\sigma}^2\right)\right).$$

Since $\widehat{\alpha} \xrightarrow{p} \alpha_0$ and given C2,

$$\begin{aligned}\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} g_{4,k}\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)\right) &= \operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} Y_t^2 Y_{t-(k+1)}^2\right) - \\ &\quad \left(\alpha_0+\beta\right)\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} Y_t^2 Y_{t-k}^2\right) - \left(\sigma_0^2\right)^2 + \left(\alpha_0+\beta\right)\left(\sigma_0^2\right)^2 \\ &= \left(\beta_0-\beta\right)\left(\alpha_0\lambda_0 + \left(\alpha_0+\beta_0\right)\eta_0\right)\left(\alpha_0+\beta_0\right)^{k-1} \\ &= E\left[g_{4,k}\left(\cdot;\ \sigma_0^2,\ \alpha_0,\ \beta\right)\right],\end{aligned} \tag{31}$$

where $\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}Y_t^2Y_{t-(k+1)}^2\right)$ is established in the proof of the Theorem. (30) and (31) grant that $T^{-1}\sum_{t=1}^{T}\overline{g}\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)\ \xrightarrow{p}\ E\left[\overline{g}\left(\cdot;\ \sigma_0^2,\ \alpha_0,\ \beta\right)\right]$ and that the only $\beta\in\Theta$ that satisfies $E\left[\overline{g}\left(\cdot;\ \sigma_0^2,\ \alpha_0,\ \beta\right)\right]=0$ is $\beta=\beta_0$. Consider the following definitions: $\overline{Q}\left(\cdot;\ \sigma_0^2,\ \alpha_0,\ \beta\right)=E\left[\overline{g}\left(\cdot;\ \sigma_0^2,\ \alpha_0,\ \beta\right)\right]'\overline{W}_0E\left[\overline{g}\left(\cdot;\ \sigma_0^2,\ \alpha_0,\ \beta\right)\right]$, $\widehat{\overline{Q}}_T\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)=\widehat{\overline{g}}_T\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)'\overline{W}$ where $\widehat{\overline{g}}_T\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)=T^{-1}\sum_{t=1}^{T}\overline{g}\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)$. Then $\widehat{\overline{Q}}_T\left(\cdot;\ \widehat{\sigma}^2,\ \widehat{\alpha},\ \beta\right)\xrightarrow{p}\overline{Q}\left(\cdot;\ \sigma_0^2,\ \alpha_0,\ \beta\right)$, which is uniquely minimized at $\beta=\beta_0$. Finally, if $\overline{W}_T=\overline{W}_T\left(\widetilde{\beta}\right)$, then (18) is the solution to (17).■

PROOF OF COROLLARY 2: If $\beta_0=0$, then

$$\widetilde{Y}_t^2=\alpha_0\widetilde{Y}_{t-1}^2+W_t,$$

and

$$\widehat{\widetilde{Y}}_t^2=\alpha_0\widehat{\widetilde{Y}}_{t-1}^2+\overline{W}_t;\quad \overline{W}_t=\left(\alpha_0-1\right)\left(\widehat{\sigma}^2-\sigma_0^2\right)+W_t.$$

For the sample moment conditions $T^{-1}\sum_{t=1}^{T}\overline{W}_t\widehat{U}_{t-1}=T^{-1}\sum_{t=1}^{T}\overline{W}_t\begin{pmatrix}X_{t-2}\\ \widehat{Z}_{t-2}\end{pmatrix}$, consider $T^{-1}\sum_{t=1}^{T}\overline{W}_tY_{t-k}$ and $T^{-1}\sum_{t=1}^{T}\overline{W}_t\widehat{\widetilde{Y}}_{t-k}^2$ for $k\geq 1$.

$$\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}\overline{W}_tY_{t-k}\right)=\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}W_tY_{t-k}\right)=0$$

by C1, C2, and C5.

$$\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}\overline{W}_t\widehat{\widetilde{Y}}_{t-k}^2\right)=\mathrm{p}\lim\left(T^{-1}\sum_{t=1}^{T}W_t\widetilde{Y}_{t-k}^2\right)=0$$

by C2, C3, and C6.

$$\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} W_t Y_{t-k}\right) = \operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\left(\widetilde{Y}_t^2-\alpha\,\widetilde{Y}_{t-1}^2\right)\,Y_{t-k}\right) \tag{32}$$
$$= \operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\widetilde{Y}_t^2 Y_{t-k}\right)-\alpha\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\widetilde{Y}_{t-1}^2 Y_{t-k}\right)$$
$$= \alpha_0^{k-1}\gamma_0\left(\alpha_0-\alpha\right)$$
$$= E\left[\left(\widetilde{Y}_t^2-\alpha\,\widetilde{Y}_{t-1}^2\right)\,Y_{t-k}\right]$$

where the third equality follows from (27).

$$\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T} W_t \widetilde{Y}_{t-k}^2\right) = \operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\left(\widetilde{Y}_t^2-\alpha\,\widetilde{Y}_{t-1}^2\right)\widetilde{Y}_{t-k}^2\right) \tag{33}$$
$$= \operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\widetilde{Y}_t^2\widetilde{Y}_{t-k}^2\right)-\alpha\operatorname{p\,lim}\left(T^{-1}\sum_{t=1}^{T}\widetilde{Y}_{t-1}^2\widetilde{Y}_{t-k}^2\right)$$
$$= \alpha_0^{k-1}\left(1-\alpha_0^2\right)^{-1}\lambda_0\left(\alpha_0-\alpha\right)$$
$$= E\left[\left(\widetilde{Y}_t^2-\alpha\,\widetilde{Y}_{t-1}^2\right)\widetilde{Y}_{t-k}^2\right]$$

where the third equality follows from (28). The only $\alpha\in\Theta$ that sets (32) and (33) to zero is $\alpha=\alpha_0$. Let $\widehat{g}_T\left(\cdot;\ \widehat{\sigma}^2,\ \alpha\right)=T^{-1}\sum_{t=1}^{T}\left(\widehat{\widetilde{Y}}_t^2-\alpha\,\widehat{\widetilde{Y}}_{t-1}^2\right)\widehat{U}_{t-1}$. Then,

$$\widehat{g}_T\left(\cdot;\ \widehat{\sigma}^2,\ \alpha\right)'\Omega_T\widehat{g}_T\left(\cdot;\ \widehat{\sigma}^2,\ \alpha\right)\xrightarrow{p} E\left[\left(\widetilde{Y}_t^2-\alpha\,\widetilde{Y}_{t-1}^2\right)U_{t-1}\right]'\Omega_0 E\left[\left(\widetilde{Y}_t^2-\alpha\,\widetilde{Y}_{t-1}^2\right)U_{t-1}\right],$$

which is uniquely minimized at $\alpha=\alpha_0$. Finally, if $\Omega_T=\Omega_T\left(\widetilde{\alpha}\right)$, then (20) is the solution to (19).■

References

[1] Anderson, T.G. and T. Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, Journal of Empirical Finance, 4, 115-158.

[2] Andrews, D.W.K., 1988, Laws of large numbers for dependent non-identically distributed random variables, Econometric Theory, 4, 458-467.

[3] Bodurtha, J.N. and N.C. Mark, 1991, Testing the CAPM with time-varying risks and returns, Journal of Finance, 46, 1485-1505.

[4] Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 307–327.

[5] Brown, B.W. and W.K. Newey, 2002, Generalized method of moments, efficient bootstrapping, and improved inference, Journal of Business and Economic Statistics, 20, 507-571.

[6] Carlstein, E., 1986, The use of subseries methods for estimating the variance of a general statistic from a stationary time series, Annals of Statistics, 14, 1171-1179.

[7] Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of Econometrics, 18, 5-46.

[8] Cragg, J.G., 1983, More efficient estimation in the presence of heteroskedasticity of unknown form, Econometrica, 51, 751-764.

[9] Donald, S.G., G. Imbens and W.K Newey, 2008, Choosing the number of moments in conditional moment restriction models, unpublished manuscript.

[10] Drost, F.C. and T.E. Nijman, 1993, Temporal aggregation of GARCH processes, Econometrica, 61, 909-927.

[11] Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, Econometrica, 50, 987-1008.

[12] Engle, R.F., and J. Mezrich, 1996, GARCH for groups, Risk, 9, 36-40.

[13] Escanciano, J.C., 2009, Quasi-maximum likelihood estimation of semi-strong GARCH models, Econometric Theory, 25, 561-570.

[14] Francq, C., L. Horath and J.M. Zakoian, 2009, Merits and drawbacks of variance targeting in GARCH models, NBER-NSF Time Series Conference proceedings.

[15] Guo, B. and P.C.B Phillips, 2001, Efficient estimation of second moment parameters in ARCH models, unpublished manuscript.

[16] Hafner, C.M., 2003, Fourth moment structure of multivariate GARCH models, Journal of Financial Econometrics, 1, 26-54.

[17] Hall, P. and J.L. Horowitz, 1996, Bootstrap critical values for tests based on generalized-method-of-moments estimators, Econometrica, 64, 891-916.

[18] Han, C. and P.C.B. Phillips, 2006, GMM with many moment conditions, Econometrica, 74, 147-192.

[19] Hansen, B., 1994, Autoregressive conditional density estimation, International Economic Review, 35, 705-730.

[20] Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, Econometrica, 50, 1029-1054.

[21] Hansen, L.P., J. Heaton and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, Journal of Business and Economic Statistics, 14, 262-280.

[22] Hansen, P.R. and A. Lunde, 2005, A forecast comparison of volatility models: does anything beat a GARCH(1,1)?, Journal of Applied Econometrics, 20, 873-889.

[23] Harvey, C. and A. Siddique, 1999, Autoregressive conditional skewness, Journal of Financial and Quantitative Analysis, 34, 465-487.

[24] Harvey, C. and A. Siddique. 2000, Conditional skewness in asset pricing tests, Journal of Finance, 55, 1263-1296.

[25] Kendall, M., 1938, A new measure of rank correlation, Biometrica, 30, 81-89.

[26] Lee, S.W, B.E. Hansen, 1994, Asymptotic theory for the GARCH(1,1) qausi-maximum likelihood estimator, 10, 29-52.

[27] Lumsdaine, R.L., 1996, Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models, Econometrica, 64, 575-596.

[28] Mark, N.C, 1988, Time-varying betas and risk premia in the pricing of forward foreign exchange contracts, Journal of Financial Economics, 22, 335-354.

[29] Meddahi, N. and E. Renault, 1997, Quadratic m-estimators for ARCH-type processes, Université de Montreal Département de sciences économiques 9814.

[30] Newey, W.K. and D.G. Steigerwald, 1997, Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models, Econometrica, 65, 587-599.

[31] Newey, W.K and F. Windmeijer, 2005, GMM with many weak moment conditions, unpublished manuscript.

[32] Prono, T., 2006, GARCH-based identification of endogenous regressors, Boston College Dissertations and Theses. Paper AAI3221276.

[33] Rich, R.W., J. Raymond and J.S. Butler, 1991, Generalized instrumental variables estimation of autoregressive conditional heteroskedastic models, Economics Letters, 35, 179-185.

[34] Skoglund, J., 2001, A simple efficient GMM estimator of GARCH models, unpublished manuscript.

[35] Smith, R.J., 1997, Alternative semi-parametric likelihood approaches to generalized method of moments estimation, The Economic Journal, 107, 503-519.

[36] Spearman, C., 1904, The proof and measurement of association between two things, American Journal of Psychology, 15, 72-101.

[37] Taskinen, S., H. Oja and R.H. Randles, 2005, Multivariate nonparametric tests of independence, Journal of the American Statistical Association, 100, 916-925

[38] Weiss, A.A., 1986, Asymptotic theory for ARCH models: estimation and testing, Econometric Theory, 2, 107-131.

[39] White, H., 1982, Instrumental variables regression with independent observations, Econometrica, 50, 483-499.

[40] Zadrozny, P.A., 2005, Necessary and sufficient restrictions for existence of a unique fourth moment of a univariate GARCH(p,q) process, CESIFO Working Paper No. 1505.

TABLE 1

The GARCH(1,1) Model

Parameter	True Value	Estimator	Med. Bias	MDAE	Dec. Range	SD
σ	1	QMLE	-0.003	0.047	0.170	0.066
		CUE	0.003	0.035	0.161	0.072
		GMM	0.028	0.050	0.215	0.100
		OLS/TSLS/CUE	-0.005	0.044	0.170	0.067
		OLS/TSLS/GMM	-0.006	0.047	0.170	0.065
α	0.05	QMLE	-0.001	0.007	0.029	0.011
		CUE	-0.001	0.004	0.021	0.013
		GMM	-0.001	0.015	0.061	0.026
		OLS/TSLS/CUE	-0.001	0.019	0.075	0.032
		OLS/TSLS/GMM	-0.002	0.020	0.077	0.031
β	0.90	QMLE	-0.001	0.015	0.058	0.024
		CUE	-0.029	0.038	0.167	0.100
		GMM	-0.058	0.063	0.217	0.100
		OLS/TSLS/CUE	-0.029	0.047	0.250	0.147
		OLS/TSLS/GMM	-0.053	0.058	0.246	0.117

Notes: Simulations are conducted using 5,000 observations across 1,000 trials. QMLE is the quasi-maximum likelihood estimator of ϑ_0. CUE is the continous-updating estimator of θ_0. GMM is the traditional two-step GMM estimator of θ_0. OLS/TSLS/CUE is the ordinary least squares estimator of σ_0^2, the two-step least squares estimator of α_0, and the continuous-updating estimator of β_0. OLS/TSLS/GMM is the ordinary least squares estimator of σ_0^2, the two-step least squares estimator of α_0, and the traditional two-step GMM estimator of β_0. For the continuous-updating and two-step GMM estimators, the number of lagged values is $K = 10$. Med. Bias is the median bias with respect to the true parameter value. MDAE is the median absolute error with respect to the true parameter value. Dec. Range is the decile range, which is the difference between the 90th and the 10th percentiles of the parameter estimates. SD is the standard deviation of the parameter estimates.

TABLE 2

The GARCH(1,1) Model

Parameter	True Value	T	Med. Bias	MDAE	Dec. Range	SD
σ	1	10K	-0.002	0.031	0.114	0.046
		20K	0.000	0.023	0.081	0.032
		40K	-0.001	0.016	0.061	0.024
α	0.05	10K	-0.004	0.016	0.058	0.023
		20K	-0.002	0.011	0.042	0.017
		40K	-0.001	0.008	0.030	0.012
β	0.90	10K	-0.021	0.034	0.145	0.064
		20K	-0.009	0.025	0.096	0.044
		40K	-0.004	0.019	0.071	0.036

Notes: Simulations are conducted across 1,000 trials. Results are reported for the OLS/TSLS/GMM estimation approach, where $\widehat{\sigma}^2$ is obtained via ordinary least squares, $\widehat{\alpha}$ via two-stage least squares, and $\widehat{\beta}$ via traditional two-step GMM. The number of lagged values used is $K = 10$. T is the number of observations per simulation trial. Med. Bias is the median bias with respect to the true parameter value. MDAE is the median absolute error with respect to the true parameter value. Dec. Range is the decile range, which is the difference between the 90th and the 10th percentiles of the parameter estimates. SD is the standard deviation of the parameter estimates.

TABLE 3

The ARCH(1) Model

Parameter	True Value	Estimator	Med. Bias	MDAE	Dec. Range	SD
σ	1	QMLE	-0.002	0.029	0.104	0.041
		CUE	-0.020	0.043	0.154	0.064
		GMM	-0.011	0.044	0.159	0.064
		OLS/CUE				
		OLS/GMM	-0.005	0.029	0.110	0.044
		OLS/IV	-0.005	0.029	0.110	0.044
		OLS/OLS	-0.004	0.029	0.108	0.043
α	0.20	QMLE	-0.005	0.023	0.087	0.034
		CUE	-0.020	0.036	0.124	0.058
		GMM	-0.027	0.040	0.131	0.061
		OLS/CUE				
		OLS/GMM	-0.030	0.041	0.115	0.054
		OLS/IV	-0.029	0.039	0.114	0.053
		OLS/OLS	-0.031	0.040	0.116	0.055

Notes: Simulations are conducted using 5,000 observations across 1,000 trials. QMLE is the quasi-maximum likelihood estimator of ϑ_0. CUE is the continous-updating estimator of θ_0. GMM is the traditional two-step GMM estimator of θ_0. OLS/CUE is the ordinary least squares estimator of σ_0^2 and the continuous-updating estimator of α_0. OLS/GMM is the ordinary least squares estimator of σ_0^2 and the traditional two-step GMM estimator of α_0. OLS/IV is the ordinary least squares estimator of σ_0^2 and the instrumental variables estimator of α_0. OLS/OLS is the ordinary least squares estimator of σ_0^2 and the ordinary least squares estimator of α_0.For the continuous-updating and two-step GMM estimators, the number of lagged values is $K = 10$. Med. Bias is the median bias with respect to the true parameter value. MDAE is the median absolute error with respect to the true parameter value. Dec. Range is the decile range, which is the difference between the 90th and the 10th percentiles of the parameter estimates. SD is the standard deviation of the parameter estimates.

TABLE 4

The GARCH(1,1) Model

Parameter	True Value	Estimator	Med. Bias	MDAE	Dec. Range	SD
σ	1	QMLE	-0.013	0.108	0.422	0.175
		CUE	-0.052	0.104	0.361	0.176
α	0.15	QMLE	-0.001	0.013	0.050	0.020
		CUE	-0.017	0.034	0.129	0.060
β	0.80	QMLE	0.000	0.014	0.059	0.023
		CUE	-0.045	0.063	0.301	0.156

Notes: Simulations are conducted using 5,000 observations across 1,000 trials. QMLE is the quasi-maximum likelihood estimator of ϑ_0. CUE is the continous-updating estimator of θ_0 based on the sample moments $T^{-1}\sum_{t=1}^{T} g_1(\cdot;\ \theta) - T^{-1}\sum_{t=1}^{T} g_3(\cdot;\ \theta)$ from (13). The number of lagged values used for the continuous-updating estimator is $K = 10$. Med. Bias is the median bias with respect to the true parameter value. MDAE is the median absolute error with respect to the true parameter value. Dec. Range is the decile range, which is the difference between the 90th and the 10th percentiles of the parameter estimates. SD is the standard deviation of the parameter estimates.

TABLE 5

The ARCH(1) Model

Parameter	True Value	Estimator	Med. Bias	MDAE	Dec. Range	SD
σ	1	QMLE	-0.007	0.044	0.159	0.064
		CUE	-0.020	0.048	0.176	0.075
		OLS/OLS	-0.016	0.048	0.181	0.109
α	0.41	QMLE	-0.007	0.029	0.109	0.043
		CUE	-0.061	0.072	0.187	0.077
		OLS/OLS	-0.101	0.106	0.210	0.088

Notes: Simulations are conducted using 5,000 observations across 1,000 trials. QMLE is the quasi-maximum likelihood estimator of ϑ_0. CUE is the continous-updating estimator of θ_0 based on the sample moments $T^{-1}\sum_{t=1}^{T} g_1(\cdot;\ \theta) - T^{-1}\sum_{t=1}^{T} g_3(\cdot;\ \theta)$ from (13). The number of lagged values used for the continuous-updating estimator is $K = 10$. Med. Bias is the median bias with respect to the true parameter value. MDAE is the median absolute error with respect to the true parameter value. Dec. Range is the decile range, which is the difference between the 90th and the 10th percentiles of the parameter estimates. SD is the standard deviation of the parameter estimates.